

DOCUMENT RESUME

ED 185 117

TM 800 151

AUTHOR Steele, Maryann E.
TITLE Effects on Predictive Ability of the Linear versus Location Models in Discriminant Analysis.
PUB DATE Oct 79
NOTE 28p.; Paper presented at the Annual Meeting of the Southern Association for Institutional Research (Orlando, FL, October 1979).
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Freshmen; *Discriminant Analysis; Dropouts; Higher Education; Mathematical Models; Nonparametric Statistics; *Predictive Measurement; Predictor Variables; Research Design
IDENTIFIERS Binary Variables; Fisher Discriminant Function; Mahalanobis Distance Function

ABSTRACT

The Mahalanobis distance model was compared with the linear discriminant function model and found to provide very similar results, even when a number of the variables were binary. A group of college freshmen were categorized into two groups: 116 "leavers," students who did not return for the second year, and 269 "returners." Data from the questionnaire Survey of Entering Freshmen was used for the predictor variables, including 34 continuous variables and 7 binary variables, including sex. Regardless of whether 7, 5, 3, or no binary variables were included, the two procedures give identical results. However, when the sample was divided into four groups, using sex as another category, minor differences in results from the two methods did appear. Adding the binary variables to the predictor variables always increased the effectiveness of the analysis, even though they violated the assumption of linearity. (Author/CTM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED185117

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

EFFECTS ON PREDICTIVE ABILITY OF THE LINEAR VERSUS

LOCATION MODELS IN DISCRIMINANT ANALYSIS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Maryann E. Steele

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Maryann E. Steele
Research Associate,
Office of Planning and
Institutional Research
St. Edward's University
3001 S. Congress Ave.
Austin, Texas 78704
(512) 444-2621

Paper presented at the 1979 Annual Meeting of the Southern Association for Institutional Research; Orlando, Florida; October, 1979.

IM800 151

ABSTRACT

The Wilk's linear model and Mahalanobis distance model were tested for predictive ability in a discriminant analysis framework, varying the number of binary variables in the analysis. Violating the assumptions of the linear model by including binary data did not reduce its predictive ability when compared to the distance model. Including more binary data increased the predictive ability in a four group design supporting the general prediction rule - the more variables included in the analysis the better.

EFFECTS ON PREDICTIVE ABILITY OF THE LINEAR VERSUS LOCATION MODELS IN DISCRIMINANT ANALYSIS

Research in higher education is progressing from description to prediction. One method of prediction widely used in the social sciences is discriminant analysis. A quick survey of Research in Higher Education from the past six years has revealed 15 to 20 articles using this method.

Green and Tull (1970) list among the objectives of discriminant analysis- "to find linear combinations of the predictor variables that enable the analyst to represent the groups by maximizing among-group relative to within-group separations." Linear analyses are proper only with interval- or ratio-scaled data (Blalock, 1979). In many cases in higher education the data are not interval- or ratio-scaled (i.e., continuous) but in fact are nominal-scaled (i.e., binary). The most common example of this being sex. The question then is, If we are analyzing binary data with continuous variable techniques, how much predictive ability do we lose? Krzanowski (1975) studied this problem by analyzing the same sets of data with both a linear and a distance (location) model. The linear model was Fisher's linear discriminant function and the distance model used the Mahalanobis method, with either three or four binary variables and up to eight continuous variables included for each method. Krzanowski concluded that the location model gave satisfactory prediction results: equal to the linear model on three data sets and much better results on two data sets.

This study compares the two models, linear versus location, in a discriminant analysis framework varying the number of binary variables from

zero to seven, along with 31 continuous variables. In the Statistical Package for the Social Sciences' (Nie, Hull, Jenkins, Steinbrenner and Bent, 1970) section on Discriminant Analysis among the methods listed, Wilk's is a linear model and Mahalanobis is a distance model. The SPSS package will be used because of its wide use in colleges and universities across the country. Prediction results (percent of known cases correctly classified) will be compared for zero, three, five or seven binary variables in both the linear and location models for two-group and four-group designs analyzing attrition/retention data.

SAMPLE CHARACTERISTICS

Seventy-six percent of the entering fall 1978 freshman (n=385) at St. Edward's University completed the Survey of Entering Freshmen, which provided information on family background, high school experiences, expectations, career plans, reasons for attending SEU, self-evaluations of abilities, and objectives. At the beginning of the fall 1979 semester, the group was divided into leavers and returners. The two students who missed the spring 1979 semester but were enrolled for fall 1979 (technically stopouts) were included in the returner group as were the students who were enrolled for three consecutive semesters. Leavers were students who had dropped out for the spring 1979 or fall 1979 semesters and had not re-enrolled at SEU. The leavers were not subdivided according to reason for leaving (i.e., academic dismissals versus voluntary withdrawals) because of the limited number of leavers (n=116) in the sample.

In looking at demographic data describing the leavers, returners and total freshman class for 1978, the groups were representative according to

Table 1 about here

sex and citizenship, fairly representative for ethnicity, but different for religious preference.

The two groups were divided equally according to sex: forty-two percent of the leavers were female while forty-seven percent of the returners were female. Forty-five percent of the total freshmen population for that year were female. The sample was also representative of the population in reference to citizenship. For fall 1978 freshmen, eleven percent were non-resident aliens. Fifteen percent of the leavers and fourteen percent of the returners reported that they were not citizens of the U.S. There were some differences between groups according to religious preference. Since SEU is a Catholic university and the study was concerned with binary variables, the responses for religious preference were Catholic and other. For the leavers, seventy-seven percent responded that they were Catholic, while for the returners, sixty-five percent said they were Catholic. For the total number of students enrolled for fall 1978, of those stating a religious preference, fifty-seven percent said they were Catholic (breakdowns by level are not available for religious preference). In reference to ethnicity, fifty-four percent of the freshmen class stated that they were white. For the leavers, forty-five percent were; forty-nine percent of the returners were white.

STUDY DESIGN

From the fall 1978 Survey of Entering Freshmen, there were 127 data points describing the freshmen. Sixty of these variables were attitudinal variables measuring reasons for enrolling, self-ratings on abilities and

importance of achieving general objectives.

A principal components analysis with varimax rotation was performed on the 60 items. Table 2 shows the nine factors that resulted. Twenty-

Table 2 about here

five of the original 60 items are included in these nine factors (factor loadings are given in Table 2). These 25 items were retained individually as variables.

From the remaining data points, nine continuously scored variables were selected for inclusion in the variable list. Seven variables were

Table 3 about here

selected as binary variables. Five of these variables were naturally dichotomous and two were dichotomized by the researcher (indicated in Table 3). Therefore, there were 41 variables total in the "7 binary" run, 39 variables in the "5 binary" run, 37 in the "3 binary" run and 34 variables in the "zero binary" run. The data was first run with two discriminant groups - leavers and returners, referred to hereafter as the Two Group Design. The discriminant analysis was then conducted with four discriminant groups - male leavers, female leavers, male returners and female returners. This is referred to as the Four Group Design. Sex was picked to be used as a discriminating category for the Four Group Design because it is a naturally dichotomous variable, available from almost any study. In the Two Group Design analysis, sex was neither the best binary predictor nor the worst. Since sex was removed from the binary variable list for the Four Group Design, the variable "kind of high school attended," scored public/other was added to maintain seven

binary variables with three binary variables now being artificially dichotomous (see Table 3). On subsequent analyses when the number of binary variables was reduced, the method of reduction was to eliminate the two binary variables that appeared lowest in the list of variables in order of entry into the prediction equation. All binary variables were eliminated for the "zero binary" run.

Because of space limitations, not all of the discriminant functions will be presented in tables nor have tables of the variable means and standard deviations by group been included. However, these are available from the author upon request. The first ten variables to enter into the prediction equation were selected by the author to be presented in tables. This was done in the interest of limited space and it is believed that from the first ten variables the reader will be able to discern any patterns that may be present.

ANALYSIS AND RESULTS

Two Group Design

A discriminant analysis of the four levels of binary variables (seven, five, three or zero) was run using the Wilk's method and then the Mahalanobis method, producing eight discriminant analyses distinguishing between two groups - leavers and returners. There was no difference found in the percent of known cases correctly classified between the Wilk's method and the

Table 4 about here

Mahalanobis method (nor in the group centroids, Table 5). Since there was also no difference found between the two methods in regard to the order the

variables entered into the prediction equation or the standardized coefficients (Tables 6 thru 9) only one set of variables and standardized coefficients are given for each of the four levels of binary variables, representing both the Wilk's and Mahalanobis methods. Comparisons will then be restricted to levels of binary variables. It was this failure to obtain any differences between the two methods that lead to conducting the Four Group Design. One possible explanation for there being no difference between the two methods (in predictive power, order entered into the equation or standardized coefficients) was that in SPSS the Mahalanobis method is theoretically identical to the Wilk's method when there are only two discriminant groups.

The discriminant analysis of the "7 binary" run produced one function which had a canonical correlation with group membership of .48 and a Wilk's lambda of .77 ($p < .07$). The centroid for leavers was .81 and for returners was -.38.

Table 6 about here

Table 6 lists the standardized discriminant function coefficients in stepwise order according to the ability of the variables to minimize Wilk's lambda and to maximize the distance between the closest groups (Mahalanobis). Of the first ten predictor variables, four of them were binary (out of seven possible). These four variables, religious preference, job for the fall, SEU first choice and sex were also in the list of the first ten predictors for the "5 binary" run. In this case, four out of five binary variables

Table 7 about here

entered into the prediction equation within the first ten steps. The "5 binary" run produced one function which had a canonical correlation with group membership of .48 and a Wilk's lambda of .77 ($p < .05$). Group centroids were: leave, .81 and return, -.37. For the "3 binary" run,

Table 8 about here

all three binary variables entered within the first ten steps. The order of the binary variables was maintained: religious preference, job for the fall and SEU first choice. The "3 binary" run produced one function which had a canonical correlation with group membership of .47 and a Wilk's lambda of .78 ($p < .023$). The group centroid for leavers was .79 and for returners -.36. For the "zero binary" run, one function was produced which had a canonical correlation with group membership of .42 with a Wilk's lambda of .83 (not significant). The group centroids were .67 for leavers and -.31 for returners. Since there are no binary variables to look at,

Table 9 about here

comparisons of the non-binary variables with the non-binary variables from the other binary level runs shows us that the variable "self rating on

Table 10 about here

writing ability" was the first non-binary variable to enter into the equation for each level. Six non-binary variables will describe the first five non-binary steps for all four levels. There is a consistency among the variables for the four levels. The binary variables were important predictor variables - they entered first and second into the prediction equation

for the three runs containing binary variables. However, although the binary variables entered into the prediction equations, they did not have any effect on the prediction results. The percent of total known cases correctly classified for the "7", "5", "3" and "zero binary" runs were 76.3%, 75.7%, 74.3% and 75.9% respectively (Table 4).

Four Group Design

A discriminant analysis of the four levels of binary variables was run using the Wilk's and Mahalanobis methods producing eight discriminant analyses distinguishing between four groups - male leavers, female leavers, male returners and female returners. There were no differences found in the percent of known cases correctly classified between the two methods for the "7", "3" or "zero binary" runs (see Table 4). Slight differences in predictive ability were found in the "5 binary" run. These differences for the "5 binary" run were also seen for the group centroids (see Table 5). Since these differences were only on one of the four runs, the author is lead to believe that this occurred by chance, as opposed to real differences. There were differences, however, between the two methods concerning the order the variables entered into the prediction equations.

On the "7 binary" run for both Wilk's and Mahalanobis, the four groups produced three functions that accounted respectively for 46, 30 and 19 percent of the discriminating power and had canonical correlations with group membership of .56, .48 and .40. Wilk's lambda was .44 ($p < .008$); when it was placed in the first function the lambda was .64 (not significant), and when more discriminating power was removed by placing it in the second function, the lambda was .84 (not significant). The group centroids distinguished more on sex than for leave and return. The centroids were:

male leave, -1.029; female leave, .276; male return, -.231; and female return, .779. Concerning the order the variables entered into the pre-

Table 11 about here

diction equation, for the Wilk's method, four out of the first ten variables were binary variables. With the Mahalanobis method, three out of the first ten variables were binary variables. The three binary variables that entered for the Mahalanobis method, SEU first choice, job for the fall and religious preference were the first three binary variables entered for the Wilk's method.

The "5 binary" run was where the differences were found. For the Wilk's method, the four groups produced three functions which accounted respectively for 46, 33 and 20 percent of the discriminating power and had canonical correlations with group memberships of .56, .50 and .40. Wilk's lambda was .43 ($p < .004$); in the first function it was .63 (not significant); and when placed in the second function, .83 (not significant). The group centroids grouped on sex: male leavers, -1.047; female leavers, .106; male returners, -.092; and female returners, .837.

For the Mahalanobis method, the three functions accounted for 44, 30 and 19 percent of the discriminating power and had canonical correlations with group memberships of .55, .48 and .40. Wilk's lambda was .45 ($p < .004$); .65 (not significant) for the first function and for the second function, .84 (not significant). Group centroids were: male leavers, -1.012; female leavers, -.197; male returners, -.213; and female returners, .780. Concerning the order the variables entered into the prediction equation, for the Wilk's method, four out of the first ten variables were binary. For the

Table 12 about here

Mahalanobis method, two out of the first ten variables were binary. These two variables, job for the fall and religious preference, were the first two binary variables to enter into the equation for the Wilk's method.

For the "3 binary" run, once again the descriptive data for the discriminant analysis was the same for the Wilk's and Mahalanobis methods. The three functions accounted for 40, 27 and 17 percent of the discriminating power and had canonical correlations of .53, .46 and .38. Wilk's lambda was .48 ($p < .008$). When placed in the first function it was .67 (not significant) and in the second function .85 (not significant). Group centroids were: male leavers, -.831; female leavers, .346; male returners, -.351; and female returners, .754.

Table 13 about here

For both methods, all three binary variables were within the first ten variables that entered into the prediction equation, in the same order.

For the "zero binary" run, identical results were obtained for the Wilk's and Mahalanobis methods. The three functions accounted for 38, 18 and 12 percent of the discriminating power and had canonical correlations with group membership of .52, .39 and .32. Wilk's lambda was .55 ($p < .075$). For the first function it was .76 and for the second .90 (neither significant). The group centroids were: male leavers, -.899; female leavers, .409; male returners, -.260; and female returners, .695.

For the Wilk's method with no binary variables, the first three variables that entered into the prediction equation were identical to the first

three non-binary variables from each of the other three runs (see Table 10). Nine non-binary variables will describe the first five non-binary steps for all four levels. For the Mahalanobis method the first non-binary variable was the same for all four runs, but then it takes 12 non-binary variables to describe the first five non-binary steps for all four levels (see Table 10). The percent of total known cases correctly classified for the "7", "5", "3" and "zero binary" runs were respectively 63.4%, 61.0%, 61.5% and 58.8% for Wilk's (see Table 4). The predictive ability for Mahalanobis was 63.4%, 62.2%, 61.5% and 58.8%.

DISCUSSION

Prediction is rapidly becoming one of the most important tasks of institutional researchers. This study examined two methods used in the prediction tool discriminant analysis - the Wilk's method and the Mahalanobis method. Since Wilk's is a linear model and Mahalanobis is a distance or location model, they are based on different assumptions. Specific to this instance, the linear model assumes a continuously - scored data base. But, not all data obtained are continuous variables; many are binary. The question stated in the introduction was, If we are analyzing binary variables with continuous variable techniques, how much predictive ability do we lose? The answer seems to be "none". There were no differences between the two methods in regard to predictive ability, group centroids, order the variables entered into the equation, discriminant function coefficients, canonical correlations or Wilk's lambda for the two group design with either seven, five, three or zero binary variables. This failure to obtain differences between the methods was attributed to the fact that for two group discriminant analysis the

methods are theoretically identical. This lead to conducting the same analyses with four groups defining the discriminant analysis. Although differences were found for the order the variables entered into the equation and standardized coefficients, no differences were found for predictive ability, except for one of the four levels. It is believed by the author that this occurred by chance.

While there was no difference between the methods, another way of looking at the problem was to see how different numbers of binary variables affected the predictive ability. When violating the assumptions of a linear model, one would expect worse predictive ability with the greater the number of binary variables available for analysis. It turns out that the test is robust and the overriding rule is that of general prediction models - the more variables available the better. This study did not find worse predictive ability for seven binary variables included in the analysis as opposed to no binary variables. For the Two Group Design, there was no difference found in the predictive ability for the extremes (76.3% correctly classified for the "7 binary" run and 75.9% for the "zero binary" run. There were differences for the Four Group Design. The percent of known cases correctly classified for the "7 binary" run was 63.4%; for the "zero binary" run it was 58.8%. While none of the predictions are as strong as would be desired, there was better prediction with the "7 binary" run, suggesting that violating the assumptions does not hinder predictive ability.

When institutional researchers are trying to predict retention, or anything else, it is not only permissible, but advisable to include binary variables.

Table 1
DEMOGRAPHIC DATA

	<u>Leave</u>	<u>Return</u>	<u>Fall 1978 Freshman Population</u>
Sex			
Male	58%	53%	55%
Female	42%	47%	45%
	<u>100%</u>	<u>100%</u>	<u>100%</u>
Citizenship			
U.S. Citizen	85%	86%	89%
Non-Resident Alien	15%	14%	11%
	<u>100%</u>	<u>100%</u>	<u>100%</u>
Religious Preference			
Catholic	77%	65%	57%
Other	23%	35%	43%
	<u>100%</u>	<u>100%</u>	<u>100%*</u>
Ethnicity			
White	45%	49%	54%
Other	55%	51%	46%
	<u>100%</u>	<u>100%</u>	<u>100%</u>

*Population = total number of students enrolled for fall 1978
stating a religious preference

Table 2

VARIMAX ROTATED FACTOR LOADINGS

ATTITUDINAL VARIABLES

Factor I	Self confidence (intellectual)	
	importance for enrolling - independence offered by SEU	.73
	self rating on academic ability	.67
	self rating on reading ability	.82
	self rating on resourcefulness	.80
	self rating on self-confidence (intellectual)	.72
Factor II	self rating on writing ability	.84
	Competency	
	objectives important to achieve - to effectively express myself verbally	.92
	objectives important to achieve - to effectively express myself in writing	.80
	objectives important to achieve - to attain specific job skills that will be useful on a job	.61
Factor III	Happy	
	self rating on cheerfulness	.64
	self rating on liking other people	.85
Factor IV	self rating on optimism	.78
	Leadership	
	self rating on ability to handle stress	.84
Factor V	self rating on leadership ability	.86
	Think	
	objectives important to achieve - to clarify and articulate values	.88
Factor VI	objectives important to achieve - to reason critically and rationally	.89
	Learn	
	objectives important to achieve - to improve my chances of making more money	.89
Factor VII	objectives important to achieve - to develop the ability to read and comprehend	.85
	objectives important to achieve - to handle without difficulty basic math problems	.75
	Feelings	
Factor VIII	self rating on impulsiveness	.82
	self rating on sensitivity to criticism	.89
Factor IX	SEU	
	importance for enrolling - I have friends at SEU	.82
	importance for enrolling - it is a Catholic school	.83
	importance for enrolling - my parents wanted me to go	.66
Factor IX	Tuition	
	importance for enrolling - the ¹⁷ tuition of SEU	.84

Table 3

CONTINUOUS AND BINARY VARIABLES

Continuous variables: (Two Group and Four Group Designs)

How far away is your home from SEU?
How often do you attend church?
Father's educational level
Mother's educational level
What was your family income last year?
In what size community did you live while attending high school?
Do you have any concern about your ability to finance your college education?
What is the highest academic degree that you intend to obtain?

Binary variables: (Two Group Design)

- * Sex (male/female)
- * Religious preference (Catholic/other)
- * Ethnicity (white/other)
- Is SEU your first choice? (yes/no)
- Have you visited SEU? (yes/no)
- Do you have a job for the fall? (yes/no)
- Are you a U.S. citizen? (yes/no)

Binary variables: (Four Group Design)

- * Religious preference (Catholic/other)
- * Ethnicity (white/other)
- Is SEU your first choice? (yes/no)
- Have you visited SEU? (yes/no)
- Do you have a job for the fall? (yes/no)
- Are you a U.S. citizen? (yes/no)
- * From what kind of high school did you graduate? (public/other)

* artificially dichotomous

Table 4

PERCENT OF KNOWN CASES CORRECTLY CLASSIFIED

Two Group Design

		Wilk's	Mahalanobis
7 binary	leave	53.5%	53.5%
	return	86.9%	86.9%
	total	76.3%	76.3%
5 binary	leave	50.7%	50.7%
	return	87.1%	87.1%
	total	75.7%	75.7%
3 binary	leave	47.9%	47.9%
	return	86.5%	86.5%
	total	74.3%	74.3%
0 binary	leave	50.0%	50.0%
	return	87.8%	87.8%
	total	75.9%	75.9%

Four Group Design

7 binary	male leave	56.5%	56.5%
	female leave	52.0%	52.0%
	male return	69.2%	69.2%
	female return	65.3%	65.3%
	total	63.4%	63.4%
5 binary	male leave	56.5%	56.5%
	female leave	48.0%	52.0%
	male return	68.4%	69.6%
	female return	60.3%	61.3%
	total	61.0%	62.2%
3 binary	male leave	47.8%	47.8%
	female leave	60.0%	60.0%
	male return	68.4%	68.4%
	female return	63.2%	63.2%
	total	61.5%	61.5%
0 binary	male leave	53.2%	53.2%
	female leave	48.0%	48.0%
	male return	62.5%	62.5%
	female return	61.8%	61.8%
	total	58.8%	58.8%

Table 5
GROUP CENTROIDS

Two Group Design

		Wilk's	Mahalanobis
7 binary	leave	.809	.809
	return	-.376	-.376
5 binary	leave	.807	.807
	return	-.370	-.370
3 binary	leave	.789	.789
	return	-.361	-.361
0 binary	leave	.672	.672
	return	-.310	-.310

Four Group Design

7 binary	male leave	-1.029	-1.029
	female leave	.276	.276
	male return	-.231	-.231
	female return	.779	.779
5 binary	male leave	-1.047	-1.012
	female leave	.106	.197
	male return	-.092	-.213
	female return	.837	.780
3 binary	male leave	-.831	-.831
	female leave	.346	.346
	male return	-.351	-.351
	female return	.754	.754
0 binary	male leave	-.899	-.899
	female leave	.409	.409
	male return	-.260	-.260
	female return	.695	.695

Table 6

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

TWO GROUP DESIGN

SEVEN BINARY VARIABLES

Wilk's and Mahalanobis

<u>Step Number</u>	<u>Variables</u>	<u>standardized coefficient</u>
* 1	Religious preference (Catholic/other)	-.53
* 2	Do you have a job for the fall? (yes/no)	-.57
3	Self rating on writing ability	-.57
4	Deciding to enroll - tuition important	-.36
5	Deciding to enroll - independence important	.34
* 6	Is SEU your first choice? (yes/no)	-.37
7	Self rating on ability to handle stress	.23
* 8	Sex (male/female)	-.19
9	Important to achieve - reason critically and rationally	.24
10	Do you have any concern about your ability to finance your college education?	-.32
* Binary variables.		

Table 7

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

TWO GROUP DESIGN

FIVE BINARY VARIABLES

Wilk's and Mahalanobis

<u>Step Number</u>	<u>Variables</u>	<u>standardized coefficient</u>
* 1	Religious preference (Catholic/other)	-.53
* 2	Do you have a job for the fall? (yes/no)	-.59
3	Self rating on writing ability	-.58
* 4	Is SEU your first choice? (yes/no)	-.36
* 5	Sex (male/female)	-.20
6	Self rating on ability to handle stress	.25
7	Deciding to enroll - independence important	.35
8	Deciding to enroll - tuition important	-.35
9	Important to achieve - reason critically and rationally	.24
10	Do you have any concern about your ability to finance your college education?	-.31

* Binary variables

Table 8

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

TWO GROUP DESIGN

THREE BINARY VARIABLES

Wilk's and Mahalanobis

<u>Step Number</u>	<u>Variables</u>	<u>standardized coefficient</u>
* 1	Religious preference (Catholic/other)	-.53
* 2	Do you have a job for the fall? (yes/no)	-.56
3	Self rating on writing ability	-.58
* 4	Is SEU your first choice? (yes/no)	-.33
5	Self rating on ability to handle stress	.26
6	Deciding to enroll - tuition important	-.41
7	Deciding to enroll - independence important	.32
8	Important to achieve - reason critically and rationally	.25
9	Self rating on cheerfulness	-.23
10	Do you have any concern about your ability to finance your college education?	-.33

* Binary variables

Table 9

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

TWO GROUP DESIGN

ZERO BINARY VARIABLES

Wilk's and Mahalanobis

<u>Step Number</u>	<u>Variables</u>	<u>standardized coefficient</u>
1	Self rating on writing ability	-.66
2	Deciding to enroll - tuition important	-.54
3	Deciding to enroll - it is a Catholic school	.27
4	Self rating on ability to handle stress	.22
5	Important to achieve - reason critically and rationally	.38
6	Self rating on academic ability	-.29
7	Deciding to enroll - independence important	.35
8	How far away is your home from SEU?	-.44
9	Self rating on resourcefulness	.29
10	In what size community did you live while attending high school?	-.20

Table 10

TOP FIVE NON-BINARY VARIABLES

Two Group Design

"7 binary" run	"5 binary" run	"3 binary" run	"zero binary" run
writing ability	writing ability	writing ability	writing ability
tuition important	handle stress	handle stress	tuition important
indep. important	indep. important	tuition important	Catholic school
handle stress	tuition important	indep. important	handle stress
reason critically	reason critically	reason critically	reason critically

Four Group Design

"7 binary" run	"5 binary" run	"3 binary" run	"zero binary" run
Wilk's			
articulate values	articulate values	articulate values	articulate values
mother's education	tuition important	tuition important	mother's education
tuition important	mother's education	mother's education	tuition important
cheerfulness	impulsiveness	impulsiveness	writing ability
handle stress	attend church	attend church	Catholic school
Mahalanobis			
tuition important	tuition important	tuition important	tuition important
father's education	clarify values	clarify values	writing ability
mother's education	cheerfulness	attend church	clarify values
like other people	handle stress	like other people	cheerfulness
family income	leadership ability	cheerfulness	read & comprehend

Table 11

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

FOUR GROUP DESIGN

SEVEN BINARY VARIABLES

Wilk's

<u>Step Number</u>	<u>Variables</u>	<u>First Function</u>
1	Important to achieve - clarify and articulate values	.30
* 2	Do you have a job for the fall? (Yes/No)	.06
3	Mother's education	.43
4	Deciding to enroll - tuition important	.57
* 5	Religious preference (Catholic/Other)	.25
6	Self rating on cheerfulness	.42
7	Self rating on ability to handle stress	-.28
8	Rank in high school graduation class	-.18
* 9	Is SEU your first choice? (Yes/No)	-.04
*10	From what kind of high school did you graduate? (Public/Other)	-.33

Mahalanobis

* 1	Is SEU your first choice? (Yes/No)	-.04
2	Deciding to enroll - tuition important	.57
* 3	Do you have a job for the fall? (Yes/No)	.06
4	Father's education	-.21
5	Mother's education	.43
6	Self rating on liking other people	-.22
7	What was your family income last year?	.42
* 8	Religious preference (Catholic/Other)	.25
9	Important to achieve - clarify and articulate values	.30
10	Deciding to enroll - it is a Catholic school	-.16

*Binary variables

Table 12
STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS
FOUR GROUP DESIGN
FIVE BINARY VARIABLES

Wilk's

<u>Step Number</u>	<u>.Variables</u>	<u>First Function</u>
* 1	Do you have a job for the fall (Yes/No)	.17
2	Important to achieve - clarify and articulate values	.31
3	Deciding to enroll - tuition is important	.56
4	Mother's education	.41
* 5	Religious preference (Catholic/Other)	.28
6	Self rating on impulsiveness	.31
* 7	Is SEU your first choice? (Yes/No)	.02
8	How often do you attend church?	-.20
9	Father's education	-.22
*10	From what kind of high school did you graduate? (Public/Other)	-.36

Mahalanobis

1	Deciding to enroll - tuition is important	.58
* 2	Do you have a job for the fall? (Yes/No)	.07
* 3	Religious preference (Catholic/Other)	.28
4	Important to achieve - clarify and articulate values	.32
5	Self rating on cheerfulness	.42
6	Self rating on ability to handle stress	-.26
7	Self rating on leadership ability	-.08
8	Self rating on writing ability	.35
9	Deciding to enroll - it is a Catholic school	-.13
10	How often do you attend church?	-.19

*Binary variables

Table 13

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

FOUR GROUP DESIGN

THREE BINARY VARIABLES

Wilk's

<u>Step Number</u>	<u>Variables</u>	<u>First Function</u>
* 1	Do you have a job for the fall? (Yes/No)	.02
2	Important to achieve - articulate and clarify values	.34
* 3	Religious preference (Catholic/Other)	.23
4	Deciding to enroll - tuition important	.56
5	Mother's education	.44
6	Self rating on impulsiveness	.31
* 7	Is SEU your first choice? (Yes/No)	-.18
8	How often do you attend church?	-.26
9	Self rating on cheerfulness	.44
10	Self rating on ability to handle stress	-.30

Mahalanobis

1	Deciding to enroll - tuition important	.56
* 2	Do you have a job for the fall? (Yes/No)	.02
* 3	Religious preference (Catholic/Other)	.23
4	Important to achieve - articulate and clarify values	.34
5	How often do you attend church?	-.26
6	Self rating on liking other people	-.18
* 7	Is SEU your first choice? (Yes/No)	-.18
8	Self rating on cheerfulness	.44
9	Self rating on leadership ability	-.07
10	Self rating on ability to handle stress	-.30

*Binary variables

Table 14

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

FOUR GROUP DESIGN

ZERO BINARY VARIABLES

Wilk's

<u>Step Number</u>	<u>Variables</u>	<u>First Function Standardized Coefficients</u>
1	Important to achieve - clarify and articulate values	.39
2	Mother's education	.42
3	Deciding to enroll - tuition important	.54
4	Self rating on writing ability	.22
5	Deciding to enroll - it is a Catholic school	-.30
6	Self rating on cheerfulness	.42
7	Self rating on ability to handle stress	-.31
8	Rank in high school graduation class	-.28
9	How often do you attend church?	-.21
10	What was your family income last year?	.40

Mahalanobis

1	Deciding to enroll - tuition important	.54
2	Self rating on writing ability	.22
3	Important to achieve - clarify and articulate values	.39
4	Self rating on cheerfulness	.42
5	Important to achieve - ability to read and comprehend	-.00
6	Rank in high school graduation class	-.28
7	Self rating on ability to handle stress	-.31
8	Deciding to enroll - independence important	-.01
9	How often do you attend church?	-.21
10	Self rating on leadership ability	-.05

REFERENCES

- Blalock, H.M. Social Statistics. New York: McGraw-Hill Book Company, Revised Second Edition, 1979.
- Green, P.E. and Tull, D.S. Research for Marketing Decisions. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.
- Krzanowski, W.J. "Discrimination and classification using both binary and continuous variables," Journal of the American Statistical Association, 70, no. 352, December 1975.
- Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H. Statistical Package for the Social Sciences. New York: McGraw-Hill Book Company, 1970.